John P. Cunningham
jpc2181@columbia.edu
Xiaoou Li
xiaoou@stat.columbia.edu
Chris Mulligan
clm2186@columbia.edu

Statistical Machine Learning (W4400)
Fall 2015
https://courseworks.columbia.edu

Homework 0

Due: Tuesday 22 September 2015

Students are encouraged to work together, but homework write-ups must be done individually and must be entirely the author's own work. Homework is due at the **beginning** of the class for which it is due. **Late homework will not be accepted under any circumstances.** To receive full credit, students must thoroughly explain how they arrived at their solutions and include the following information on their homeworks: name, UNI, homework number (e.g., HW03), and class (STAT W4400). All homework must be turned in online through Courseworks in PDF format, have a .pdf extension, and be less than 4MB. If programming is part of the assignment, the code must be turned in in one or more .R files. Homeworks not adhering to these requirements will receive no credit.

Preamble

Prerequisites for this course include a previous course in statistics, elementary probability, multivariate calculus, linear algebra. This homework is designed to allow you to test your background and your ability to adhere to the above submission instructions. All questions have been designed to be solved without any numerical aid – if you are using a computer (which you are welcome to do), you may be missing the didactic purpose.

Submission (50 points)

Submit your homework according to the above instructions.

Questions (50 points)

In all of the below questions, let:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad \text{and} \qquad x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

1. (2 points) What is $B_{2,1}$?
2. (2 points) What is $A + B$?
3. (2 points) What is AB ?
4. (2 points) What is $rank(A)$?
5. (2 points) What is the largest eigenvalue of A?
6. (2 points) What is the eigenvector associated with that largest eigenvalue of A?

7. (2 points) What is $|B|$ (determinant of B)?
8. (2 points) What is $x^\top Ax$?
9. (2 points) What is $x^\top x$?
10. (2 points) What is $xx^\top$?
11. (2 points) What is $||x||_2$ (the Euclidean norm of x)?
12. (2 points) What is the gradient of $f(x) = x^\top Ax$ w.r.t. x, namely $\nabla_x f(x)$?
13. (2 points) What is the Hessian of $f(x) = x^\top Ax$ w.r.t. x, namely $\nabla_x^2 f(x)$?
14. (2 points) We say $x \in \mathbb{R}^n$. What is n?
15. (2 points) I write $y \sim \mathcal{N}(\mu, \sigma)$ to denote a Gaussian random variable with mean μ and standard deviation σ. What is $E(y^2)$?
16. (2 points) Say $y \sim \mathcal{N}(2.7, 8)$ and $w \sim \mathcal{N}(3.1, 15)$ are independent random variables. What is the distribution of $y + w$?
17. (2 points) I write $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix}\right)$ to denote a multivariate Gaussian random variable with dimension $n = 2$, and mean vector μ and covariance matrix Σ as specified. What is the normalizing constant of this distribution?
18. (2 points) I write $z \sim Bern(p)$ to denote a Bernoulli random variable with bias p. What is the support of z?
19. (2 points) I draw n times independently from $z \sim Bern(p)$. What is the distribution of the number of heads/successes k?
20. (2 points) Find x_1 that maximizes $h(x_1) = \frac{1}{3}x_1^3 - \frac{1}{2}x_1^2 - 6x_1 + \frac{27}{2}$ subject to $x_1 \in [-4, 4]$.
21. (2 points) Find the minimum value of $h(x_1)$ subject to the constraint that $x_1 \in [-4, 4]$.
22. (2 points) Let $\tilde{h}(x_1) = \frac{1}{Z}h(x_1)$; find Z such that $\tilde{h}(x_1)$ has $\int_0^1 \tilde{h}(x_1)dx_1 = 1$.
23. (2 points) Let $b(x) = x_1 x_2^3$. Find $\int_{\mathcal{A}} b(x)dx$ for $\mathcal{A} = [0, 3] \times [0, 2]$.
24. (2 points) Let $c(x) = x_1 + \sqrt{3}x_2$; find x that maximizes $c(x)$ subject to $x_1^2 + x_2^2 = 1$.
25. (2 points) Let $g(x) = -x_1 \log x_1 - x_2 \log x_2$; find x that maximizes $g(x)$ subject to $x_1 + x_2 = 1$.

Closing Remark

Linear algebra and optimization are huge and beautiful mathematical fields, and we will only skim the very surface. That said, matrices, vectors, and common manipulations of these objects are the tools of the trade in data science, and thus a basic facility is crucial. Regardless of your ease with the above questions, for linear algebra I recommend studying Zico Kolter's excellent and brief review (for a machine learning class): `http://cs229.stanford.edu/section/cs229-linalg.pdf`. For basic use of Lagrange multipliers, I recommend both of the following:

www.cs.iastate.edu/%7Ecs577/handouts/lagrange-multiplier.pdf?.

www.math.harvard.edu/archive/21a%5Fspring%5F09/PDF/11-08-Lagrange-Multipliers.pdf

If you feel drastically behind on all these subjects, I strongly recommend serious self-study, including something like the first 16 lectures of:

http://ocw.mit.edu/courses/mathematics/18-02-multivariable-calculus-fall-2007/index.htm .

Or perhaps the entirety of: http://projects.iq.harvard.edu/stat110 .

Successful completion of this course without much of the above background will be challenging, though not impossible.

As a grading rubric, if you have no trouble answering 60-100% of these questions, you are in the right class. If you are able to correctly answer 40-60% of the questions, you will be successful as long as you continue to work to refresh these concepts. If you score well below 40%, you will struggle in this course; you may want to reconsider taking this course without developing more background. These concepts will be reviewed in brief meaningful detail, but they are the necessary toolkit to begin this material, and thus familiarity is expected.